CLAIMS

What is claimed is:

1. A method for visual-based recognition of an object, said method comprising:

receiving depth data for at least a pixel of an image of an object, said depth data comprising information relating to a distance from a visual sensor to a portion of said object shown at said pixel;

generating a plan-view image based in part on said depth data;

extracting a plan-view template from said plan-view image; and

processing said plan-view template at a classifier, wherein said classifier is trained to make a decision according to pre-configured parameters.

2. The method as recited in Claim 1 further comprising receiving non-depth data for said pixel.

3. The method as recited in Claim 1 wherein said visual sensor determines said depth data using stereopsis based on image correspondences.

4. The method as recited in Claim 1 wherein said generating said plan-view image comprises selecting a subset of said depth data based on foreground segmentation.

5. The method as recited in Claim 1 wherein said generating said plan-view image further comprises:

generating a three-dimensional point cloud of said subset of pixels based on said depth data, wherein a point of said three-dimensional point cloud comprises a three-dimensional coordinate;

partitioning said three-dimensional point cloud into a plurality of vertically oriented bins; and

mapping at least a portion of points of said plurality of vertically oriented bins into at least one said plan-view image based on said three-dimensional coordinates, wherein said plan-view image is a two-dimensional representation of said three-dimensional point cloud comprising at least one pixel corresponding to at least one vertically oriented bin of said plurality of vertically oriented bins.

6. The method as recited in Claim 4 further comprising receiving non-depth data for said pixel, and wherein said foreground segmentation is based at least in part on said non-depth data.

7. The method as recited in Claim 5 further comprising dividing said three-dimensional point cloud into a plurality of slices, and wherein said generating said plan-view image is performed for at least one slice of said plurality of slices.

8. The method as recited in Claim 7 wherein said extracting a plan-view template from said plan-view image further comprises extracting a plan view template from at least two plan-view images corresponding to different slices of said plurality of slices, wherein said plan-view template comprises a transformation of at least a portion of said plan-view images, such that said plan-view template is processed at said classifier.

9. The method as recited in Claim 1 wherein said extracting said plan-view template from said plan-view image is based at least in part on object tracking.

10. The method as recited in Claim 1 wherein said classifier is a support vector machine.

11. The method as recited in Claim 2 wherein said plan-view image is based in part on said non-depth data.

12. The method as recited in Claim 1 wherein said object is a person.

13. The method as recited in Claim 1 wherein said plan-view image comprises a value based at least in part on an estimate of height of a portion of said object above a surface.

14. The method as recited in Claim 1 wherein said plan-view image comprises a value based at least in part on color data for a portion of said object.

15. The method as recited in Claim 1 wherein said plan-view image comprises a value based at least in part on a count of pixels obtained by said visual sensor and associated with said object.

16. The method as recited in Claim 1 wherein said plan-view template is represented in terms of a vector basis.

17. The method as recited in Claim 16 wherein said vector basis is obtained through principal component analysis (PCA).

18. The method as recited in Claim 13 further comprising performing height normalization on said plan-view template.

19. The method as recited in Claim 1 wherein said decision is to distinguish between a human and a non-human.

20. The method as recited in Claim 1 wherein said decision is to distinguish between a plurality of different human body orientations.

21. The method as recited in Claim 1 wherein said decision is to distinguish between a plurality of different human body poses.

22. The method as recited in Claim 1 wherein said decision is to distinguish between a plurality of different classes of people.

23. A visual-based recognition system comprising:

a visual sensor for capturing depth data for at least a pixel of an image of an object, said depth data comprising information relating to a distance from said visual sensor to a portion of said object visible at said pixel;

a plan-view image generator for generating a plan-view image based on said depth data;

a plan-view template generator for generating a plan-view template based on said plan-view image; and

a classifier for making a decision concerning recognition of said object, wherein said classifier is trained to make a decision according to pre-configured parameters.

24. The visual-based recognition system as recited in Claim 23 wherein said visual sensor is also for capturing non-depth data.

25. The visual-based recognition system as recited in Claim 23 wherein said visual sensor determines said depth data using stereopsis based on image correspondences.

26. The visual-based recognition system as recited in Claim 23 wherein said plan-view image generator comprises a pixel subset selector

for selecting a subset of pixels of said image, wherein said pixel subset selector is operable to select said subset of pixels based on foreground segmentation.

27. The visual-based recognition system as recited in Claim 23 wherein said classifier is a support vector machine.

28. The visual-based recognition system as recited in Claim 24 wherein said plan-view image is based in part on said non-depth data.

29. The visual-based recognition system as recited in Claim 23 wherein said plan-view image generator is operable to generate a three-dimensional point cloud based on said depth data, wherein a point of said three-dimensional point cloud comprises a three-dimensional coordinate.

30. The visual-based recognition system as recited in Claim 29 wherein said plan-view image generator is operable to divide said three-dimensional point cloud into a plurality of slices such that a plan-view image may be generated for at least one slice of said plurality of slices.

31. The visual-based recognition system as recited in Claim 30 wherein said plan-view template generator is operable to extract a plan-view template from at least two plan-view images corresponding to different slices of said plurality of slices, wherein said plan-view template comprises a transformation of at least a portion of said plan-view images, such that said plan-view template is processed at said classifier

32. A method for visual-based recognition of an object representative in an image, said method comprising:

generating a three-dimensional point cloud based on depth data for at least a pixel of an image of said object, said depth data comprising

information relating to a distance from a visual sensor to a portion of said object visible at said pixel, said three-dimensional point cloud representing a foreground surface visible to said visual sensor and wherein a pixel of said three-dimensional point cloud comprises a three-dimensional coordinate;

partitioning said three-dimensional point cloud into a plurality of vertically oriented bins;

mapping at least a portion of points of said vertically oriented bins into at least one said plan-view image based on said three-dimensional coordinates, wherein said plan-view image is a two-dimensional representation of said three-dimensional point cloud comprising at least one pixel corresponding to at least one vertically oriented bin of said plurality of vertically oriented bins; and

processing said plan-view image at a classifier, wherein said classifier is trained to make a decision according to pre-configured parameters.

33. The method as recited in Claim 32 wherein said three-dimensional point cloud and said plan-view image are also based at least in part on non-depth data.

34. The method as recited in Claim 32 wherein said visual sensor determines said depth data using stereopsis based on image correspondences.

35. The method as recited in Claim 32 further comprising extracting a plan-view template from said plan-view image, wherein said plan view template comprises a transformation of at least a portion of said plan view image, and such that said plan-view template is processed at said classifier.

36. The method as recited in Claim 32 further comprising dividing said three-dimensional point cloud of into a plurality of slices, and wherein said mapping at least a portion of points comprises mapping points within a slice of said plurality of slices of said three-dimensional point cloud into said plan-view image.

37. The method as recited in Claim 36 further comprising extracting a plan-view template from said plan-view image, wherein said plan view template comprises a transformation of at least a portion of said plan view image, such that said plan-view template is processed at said classifier.

38. The method as recited in Claim 32 wherein said classifier is a support vector machine.

39. The method as recited in Claim 32 wherein said plan-view image is generated from a subset of pixels of said image selected based on foreground segmentation.

40. The method as recited in Claim 36 further comprising extracting a plan view template from at least two plan view images corresponding to different slices of said plurality of slices, wherein said plan view template comprises a transformation of at least a portion of said plan view images, such that said plan-view template is processed at said classifier.